https://doi.org/10.56093/JISAS.V79I3.2
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 79(3) 2025 207–220

Modelling and Forecasting of Tomato Prices in Jabalpur District of India

Shanya Tiwari, Umesh Singh and Kuldeep Rajpoot

Jawaharlal Nehru Krishi Vishwavidyalaya, Jabalpur

Received 21 July 2025; Revised 08 October 2025; Accepted 14 October 2025

SUMMARY

Tomato price volatility exerts a substantial influence on food inflation and poses significant challenges for agricultural market volatility management. This study employs a hybrid SARIMA–GARCH modeling framework with Root Mean Square Error (RMSE) 0.3011, Mean Absolute Percentage Error (MAPE) 3.1754, and Mean Absolute Error (MAE) 0.2356,to forecast monthly tomato prices for a horizon of 12 months in the Jabalpur market, capturing both seasonal patterns and volatility clustering in the price series. The SARIMA$(1,1,1)(1,0,0)_{12}$ model effectively captures seasonal dynamics, yielding a RMSE of 0.385, MAPE of 3.812, and MAE of 0.286. The GARCH(1,1) component addresses conditional heteroskedasticity in the residuals. To enhance model robustness, a Box–Cox transformation is applied on the price series to stabilize the variance and mitigate heteroscedastic effects. Parameters are estimated using Maximum Likelihood (ML) and a two-stage Conditional Sum of Squares followed by ML (CSS–ML) procedure. Comparative evaluation demonstrates that the hybrid SARIMA$(1,1,1)(1,0,0)_{12}$–GARCH(1,1) model outperforms both SARIMA models with varying specifications and ARFIMA (1,0.4916,1) models in terms of predictive accuracy. The hybrid model is further validated through residual diagnostics. An exploratory extension incorporating log-transformed arrivals as exogeneous variables in a SARIMAX framework yields marginal improvements, reinforcing the preference for model parsimony. The proposed hybrid approach offers a scalable and empirically validated tool for improving short-term price forecasts, with practical implications for production planning, inventory management, and policy formulation in volatile tomato markets.

Keywords: Tomato prices; Hybrid SARIMA-GARCH; Forecasting; Box-Cox transformation; Agricultural Market Volatility; ARIMA.

1. INTRODUCTION

The dynamics of inflation, particularly in the context of developing economies like India, have significant implications for economic policy and household well-being. The high inflation period of 2021-23 has underscored the urgent need to study inflation expectation dynamics, as understanding these patterns becomes increasingly critical for economic stability (Patra *et. al.*, 2023; GOI, Ministry of Finance, Department of Economic Affairs, 2025). Food price inflation, typically more volatile than core inflation, has been a key driver of changes in headline inflation and inflation expectations. The contribution of food inflation to headline inflation has been substantial at times, accounting for 67% in November 2013, 62% in July 2016, and 75% in December 2019, although the average contribution from January 2012 to March 2022 was around 47% (RBI, 2022). Examining the broader context of food and beverage inflation, the CPI-Food and Beverages group assigns the highest weight to cereals and cereal products at 21.1, followed by milk and milk products at 14.4, and then vegetables at 13.2, with the remaining weight attributed to other edible items (Pratap *et al.,* 2021). Vegetables, which are among the most volatile components within the Food and Beverages group, exert a substantial influence on the broader trajectory of food and beverage inflation. The price inflation of vegetables is primarily driven by the prices of tomatoes, onions, and potatoes (collectively referred to as the 'TOP' commodities), which have a significant combined weighting of 36.5% in the CPI-Vegetables basket, 4.8% in the CPI-Food

Corresponding author: Kuldeep Rajpoot
E-mail address: kuldeep2rajpoot@gmail.com

and beverages basket, and 2.2% in the CPI-Combined basket (Pratap *et al.,* 2021). As per historical data, tomatoes show higher price volatility (0.21) compared to onions (0.19) and potatoes (0.10) (NABARD, 2023). The strong positive correlation between tomato prices and corresponding consumer price index measures suggests that tomato price volatility can substantially impact broader inflation metrics.

Predominantly, models employed for the prediction of the homoscedastic prices of the non-volatile agricultural commodities such as cereals and oilseeds utilize linear time series models, such as Auto Regressive (AR) and Auto Regressive Integrated Moving Average (ARIMA). The exponential decay of the autocorrelation function of these models gives them the ability to capture the short-range dependence (SRD) characteristics. SARIMA (Seasonal Auto Regressive Integrated Moving Average) model is particularly designed to manage seasonality within time series data, making it ideal for agricultural prices that are influenced by annual planting and harvesting cycles. The seasonal component of SARIMA can effectively capture periodic patterns in prices. While ARIMA doesn't explicitly model seasonality, it is effective for non-seasonal time series with trends or cycles. While capable of capturing non-linear relationships, machine learning models like ANN(Artificial Neural Network) don't inherently handle seasonality or trends without significant pre processing (e.g., adding lagged variables or using feature engineering). This makes them less directly suitable for datasets with strong seasonal patterns. As a result, they might not be the first choice for purely seasonal datasets unless combined with other methods, such as SARIMA. Traditional statistical models like ARIMA, SARIMA and ARCH (Autoregressive Conditional Heteroscedasticity)/GARCH (Generalised Autoregressive Conditional Heteroscedasticity) models perform well with small datasets because they rely on historical relationships within a time series rather than requiring large amounts of training data (Damrongkulkamjorn & Churueang, 2005) and require estimation of fewer parameters while neural networks like ANN typically require larger datasets for training to learn patterns effectively. If the available time series data is limited or has many missing values, ANNs might struggle to generalize well, leading to over fitting (Casolaro *et al.,* 2023). The sporadic and irregular nature of daily price records in agricultural commodity markets, when condensed into monthly data, might create an insufficient training sample for neural networks that fails to capture complex price dynamics, leading to poor generalization and unreliable predictions. Consequently, alternative methodologies of time series analysis or hybrid machine learning approaches are more appropriate for handling the inherent variability and sparsity of agricultural commodity pricing data. ARCH/GARCH models are designed to predict time-varying volatility, which is a common characteristic in commodity prices due to factors like weather events, policy changes, or market disruptions. GARCH models are effective at capturing such volatility clusters, while ANNs can be adapted to forecast volatility, they do not inherently model variance. Developing such models often requires additional layers and tuning, making it more complex. This study is conducted with the primary objective of determining the most suitable model for forecasting tomato prices and thereafter generating accurate tomato price forecasts in Jabalpur district so as to facilitate strategic agricultural planning and market interventions.

There exists a substantial body of work pertaining to the forecasting of agricultural commodity prices. Kumar *et al.* (2011) employed Holt-Winters and seasonal ARIMA models for onion price forecasting in Bangalore, demonstrating early application of exponential smoothing in agricultural markets. Singh *et al*. (2015) developed an econometric simulation model for forecasting in dal industries, emphasizing the importance of structural modeling in policy simulation. Paul *et al.* (2016) analyzed asymmetric price volatility using forecasting techniques for onion in Delhi, indicating growing concern for price instability. Reddy *et al.* (2018) investigated pearl millet market structures and value chains, contextualizing price forecasting within broader economic frameworks. Kitworawut and Rungreunganun (2019) applied Box-Jenkins models to corn prices, validating ARIMA based methods for stable cash crops. Rajpoot *et al.* (2021) evaluated the impact of COVID-19 lockdowns on potato and onion prices in metropolitan India, illustrating the vulnerability of supply chains to systemic shocks. Jain *et al.* (2022) proposed a context-based model selection framework, advocating adaptive models over static ones for crop price forecasting. Badal *et al.* (2022)

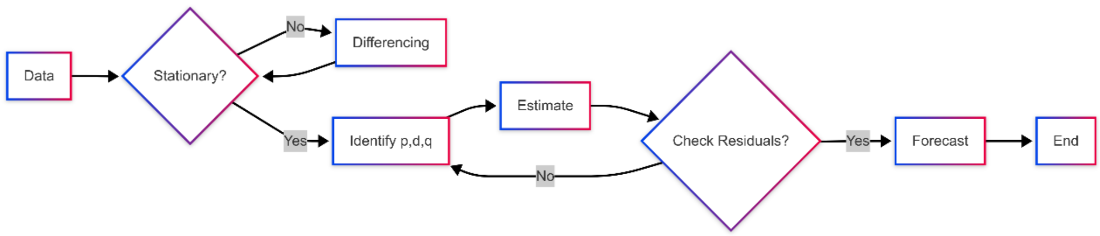

Fig. 1. Flowchart demonstrating a schematic representation of Box Jenkins methodology

used ARIMA for potato price forecasting, reinforcing its utility in periodic price prediction. Pramanik *et al.* (2023) performed a comparative evaluation of several advanced time series forecasting models, including ARIMA, GARCH, and hybrid approaches, TDNN and SVM etc. applied to volatile agricultural price data, emphasizing the relative forecasting accuracy. Sarkar *et al.* (2024) studied tomato markets in West Bengal, combining price and arrival behavior analysis to understand market dynamics. Sonvanee and Bhargav (2024) forecasted tomato prices and arrivals in Chhattisgarh's Krishi Mandis using ARIMA, showcasing its effectiveness in hyper local settings. Kumar and Priya (2024) analyzed tomato prices in Tamil Nadu's Koyambedu market, reaffirming the continued relevance of statistical forecasting. Sanjeev *et al.,* (2025) performed a Comparative Study of ARIMA, SARIMA and Hybrid (ARIMA + ANN and SARIMA + ANN) Models for Wholesale Monthly Average Price of Tomato and Onion and reported hybrid (SARIMA + ANN) technique as an effective way to improve the forecasting performance for price series of tomato and onion crop.

2. MATERIALS AND METHODOLOGY

Data: The secondary time series data analysed in this study pertains to daily tomato prices from agricultural markets (mandis) in Jabalpur district, Madhya Pradesh, India, spanning the period from June 2011 to June 2024, sourced from the AGMARKNET platform, an official initiative by the Department of Agricultural Marketing and Inspection under the Ministry of Agriculture, Government of India, accessible at https://agmarknet.gov.in/. In this study, monthly tomato time series was generated by computing the mean of daily price values and arrivals across multiple months, along with pre-processing data to handle missing data points (Knotters *et al.,* 2010).

2.1 Time Series Modelling and Box-Jenkins methodology

Time series analysis emerges as a sophisticated analytical approach to understanding sequential data characterized by temporal interdependencies. At its mathematical foundation, a time series is conceptualized as a stochastic process $\{X_t, t \in Z\}$, where each observation X_t represents a random variable indexed by discrete time. The Box-Jenkins methodology (Box & Jenkins, 1970) provides a systematic framework for time series modeling through three critical stages: model identification, estimation of parameters and diagnostic checking. The flow chart in Fig. 1 presents Box Jenkins Methodology vividly.

The identification stage involves analysing the autocorrelation function (ACF, ρ_k) and the partial autocorrelation function (PACF, ϕ_{kk}), to determine the orders of the Moving Average (MA) (q) and Autoregressive (AR) (p) components, respectively while, differencing provide order d of integration. In estimation stage Conditional Least Squares-Maximum Likelihood (CSS-ML) estimation method is used for estimation of parameters of the ARIMA model. The diagnostic checking stage, checks the adequacy of fitted models which is presented in detail below under sub-section 2.4.

2.2 Stationarity and Time Series Transformations

Stationarity represents a critical assumption in time series modelling, demanding that fundamental statistical properties remain invariant across different time periods. The Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) is a statistical method for examining stationarity in time series data through three model variations: Random Walk $\nabla X_t = \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \nabla X_{t-i} + \varepsilon_t$, Random Walk with

Drift $\nabla X_t = \alpha + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \nabla X_{t-i} + \varepsilon_t$ and Deterministic Trend Model $\nabla X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \nabla X_{t-i} + \epsilon_t$. The test's null hypothesis ($H_0 : \gamma = 0$) posits that the series is non-stationary due to the presence of a unit root, while the alternative hypothesis ($H_1 : \gamma < 0$) implies stationarity. The test statistic is calculated as,

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

comparing the estimated coefficient of lagged level term against critical values that depend on sample size, model specification, and significance level. Rejecting the null hypothesis indicates the time series is stationary, which is crucial for further statistical analysis and modelling. Key considerations include selecting an appropriate lag length, choosing the correct model specification, and utilizing information criteria like AIC and BIC for comprehensive evaluation.

The Box-Cox transformation provides a systematic approach to stabilizing variance and normalizing the data distribution,

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ ln(y), & \lambda = 0 \end{cases}$$

where λ is the transformation parameter.

2.3 Time Series Modelling

The **Seasonal Autoregressive Integrated Moving Average with exogenous regressors** provides a powerful mathematical framework for modelling complex temporal dynamics. The **SARIMAX**(p,d,q, k) $\times$ (P,D,Q)$_m$ model extends the classical autoregressive and moving average framework by incorporating external predictors, or exogenous variables, making it suitable for situations where external factors influence the time series. It is represented as

$$\Phi(B^m)(1-B^m)^D \phi(B)(1-B)^d \left(X_t - \sum_{i=1}^{k} \beta_i Z_{it}\right) = \Theta(B^m)\theta(B)\epsilon_t,$$

where Z_{it} denotes external variables such as economic indicators, weather data, or policy changes, β_i are their respective coefficients and k indicates the total number of exogenous predictors (or its lags) included in the model. The polynomial $\phi(B)$ represents the autoregressive (AR) component of degree p, capturing the dependence of X_t on its lagged values, while $\theta(B)$ represents the moving average (MA) polynomial of degree q, capturing serial correlation in the error terms. The parameter d denotes the non-seasonal differencing order required to achieve stationarity. For seasonal dynamics, m represents the seasonal period (e.g., 12 for monthly data), D represents the seasonal differencing order, and $\Phi(B^m)$ and $\Theta(B^m)$ denote the seasonal AR and MA polynomials of orders P and Q, respectively. The term ϵ_t is a zero-mean innovation process. This formulation is particularly valuable in economic forecasting, where multiple external factors may influence the target variable. When exogenous predictors are not incorporated ($Z_{it} \equiv 0$), the model reduces to a SARIMA (P, D, Q, p ,d, q)$_m$ model, and if the seasonal AR and MA polynomials are of order zero $(P = Q = 0)$, it further reduces to a standard ARIMA (p,d,q) model. The **ARFIMA** model introduces the concept of fractional differencing, suited for long-memory processes and more flexible modelling of persistence,

$$\phi(B)(1-B)^{d_f} X_t = \theta(B)\epsilon_t$$

where d_f is the fractional differencing parameter $(0 < d_f < 0.5)$. This model is particularly useful for financial time series and long-range dependent data where the autocorrelation function decays hyperbolically rather than exponentially.

Advanced time series analysis recognizes that variance itself can be time-varying, a common feature in financial time series. The Generalized ARCH (GARCH (p, q)) model (Bollerslev, 1986) captures this this dynamic by modeling the conditional variance as a function of past p squared residuals and past q conditional variances, providing a more parsimonious representation of volatility dynamics,

$$y_t = \mu + \zeta_t$$

$$\zeta_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \zeta_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where σ_t^2 is the conditional variance of heteroskedastic residuals ζ_t and $\epsilon_t \sim i.i.d.WN(0,1)$. When this conditional variance is a function of only the squared residuals and not past conditional variances (i.e. $q=0$) the model reduces to ARCH(p) model (Engle, 1982). The GARCH model is particularly effective in financial applications, capturing both volatility clustering and the leverage effect, where negative returns tend to impact volatility more than positive returns of the same magnitude.

2.4 Model Diagnostics and Selection Criteria

Model diagnostics employ various statistical tests to ensure model adequacy and validate assumptions. The **Ljung-Box** test (Ljung & Box, 1978) examines residual autocorrelation, testing the null hypothesis that residuals are independently distributed:

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\widehat{\rho_k^2}}{n-k}$$

where $\widehat{\rho_k}$ represents the sample autocorrelation at lag k, n is the total sample size of the series, $Q \sim \chi_h^2$ and h is the number of lags being tested. A significant Q-statistic indicates the presence of serial correlation in the residuals, suggesting model inadequacy.

The **ARCH LM** test investigates residual heteroskedasticity by examining whether squared residuals exhibit autocorrelation:

$$\zeta_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \zeta_{t-i}^2 + \epsilon_t$$

where ζ_t are the residuals from the model and $\epsilon_t \sim i.i.d.WN(0,1)$. This test is particularly crucial for financial time series where volatility clustering is common. The null hypothesis of no ARCH effects is rejected if the R^2 from this regression multiplied by the sample size is larger than critical value.

$$\text{LM} = nR^2, \;\; \text{LM} \sim \chi_q^2$$

To verify the distributional adequacy of the fitted models, the normality of the residuals was tested using the **Jarque–Bera (JB)** and **Shapiro–Wilk** tests, both of which evaluate departures from normality. The **Granger Causality Test** assesses predictive causality between time series, testing the null hypothesis that one variable does not provide additional predictive information, with its test statistic specified as,

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(T-k)}$$

where, RSS_R is the Residual Sum of Squares for the Restricted Model, RSS_{UR} is the Residual Sum of Squares for the Unrestricted Model, q is the number of restrictions (X variable lags), T is the total observations, k denotes the parameters in the unrestricted model. Restricted model is expressed as, $Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \epsilon_t$, while unrestricted model is expressed as, $Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j} + \eta_t$. A statistically significant F-statistic rejects the null hypothesis, indicating that lagged values of X provide additional predictive information for Y beyond its own past values.

Goodness of fit measures provide quantitative assessments of model performance. Coefficient of determination (R^2) measures the proportion of variance explained by the model:

$$R^2 = 1 - \frac{\Sigma\left(y_t - \widehat{y_t}\right)^2}{\Sigma\left(y_t - \overline{y}\right)^2}$$

Theil's U_2 statistic compares the forecasting performance against a naive model:

$$U_2 = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\widehat{y_t} - y_{t-1}}{y_{t-1}}\right)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{y_t - y_{t-1}}{y_{t-1}}\right)^2}}$$

A U_2 statistic less than 1 indicates that the model outperforms naive forecasts.

Model selection criteria balance model complexity and goodness of fit. The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-

Quinn Information Criterion (HQIC) are commonly used, which are expressed as,

$$AIC = 2k - 2\ln(L)$$

$$BIC / SC = \ln(n)k - 2\ln(L)$$

$$\text{HQIC} = -2\ln(L) + 2k\ln(\ln(n))$$

where, k is the number of parameters in the model, L is the maximum likelihood and n is sample size. Lower values of AIC, BIC, and HQIC indicate better model fit, with each criterion applying different penalties for model complexity.

Forecast accuracy metrics such as RMSE, MAE, MAPE are utilized for providing different insights into forecast accuracy.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \widehat{y}_t\right)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \widehat{y}_t\right|$$

$$\text{MAPE (\%)} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \widehat{y}_t}{y_t}\right| \times 100$$

Multistep forecasting involves predicting values for multiple future time steps, extending the forecast horizon, requiring models to account for the compounding nature of errors over successive predictions.

3. RESULTS AND DISCUSSION

The empirical analysis of the monthly average tomato arrivals and price time series data was conducted using the Box-Jenkins methodology. The time series plot of the monthly average tomato arrivals and prices in Jabalpur district is illustrated in Fig. 2 while the descriptive statistics presented in Table 1.

The Box-Cox transformation of the price series data was applied for stabilizing the variance and normalizing the distribution (Vinay *et al.,* 2024). The estimated value of the power parameter λ (−0.008279643) is very close to zero, suggesting that the transformation is nearly equivalent to a logarithmic transformation. Consequently, a logarithmic transformation of the monthly price series has been employed. Non-normality

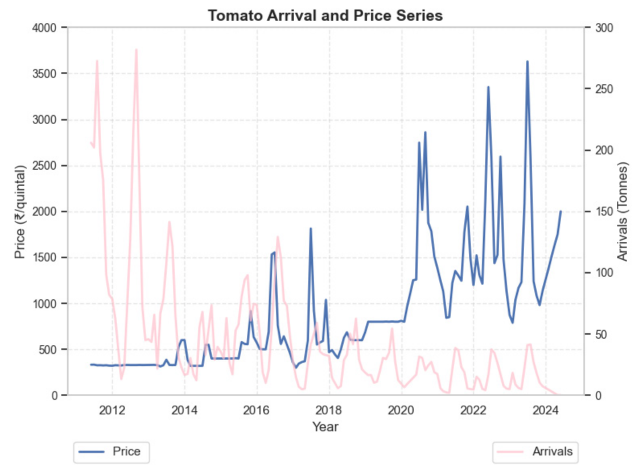

Fig. 2. Monthly Average Tomato Arrival and Price Series Plot

Table 1. Descriptive Statistics of Tomato Price and Arrival series of Jabalpur District

Statistics	Original Price Series	Log Transformed Price Series	Arrival Series
Min	299.63	5.70	0.20
1st Quartile	386.20	5.96	13.28
Median	607.14	6.41	28.81
Mean	879.59	6.56	44.08
3rd Quartile	1220.00	7.11	55.54
Max	3630.43	8.20	281.87
Standard Deviation	650.32	0.65	50.24
Coefficient of Variation (CV)	0.74	0.10	1.14
Skewness	1.67	0.44	2.42
Kurtosis	2.97	-0.87	6.64
Jarque-Bera χ^2	135.04	9.89	454.67
Jarque-Bera p-value	$< 2.2 \times 10^{-16}$***	0.007114 ***	$< 2.2 \times 10^{-16}$ ***

persists in both the original and log-transformed price series, as evidenced by the deviations observed in the Q–Q plot in Fig. 3 and the results of the Shapiro–Wilk test, which yielded very low *p*-values 4.76×10^{-13} and 4.768×10^{-07} ($p< 0.01$) for both the original and log-transformed price series with Shapiro-Wilk statistic W, 0.81 and 0.93, respectively indicating significant departures from normality. However, the higher W statistic (closer to 1) for the log-transformed series suggests that the transformation has brought the data closer to normality, although it does not achieve it. This improvement implies that the original data may follow a distribution approximating log-normality, a pattern commonly observed in price data (Limpert *et al.,* 2001).

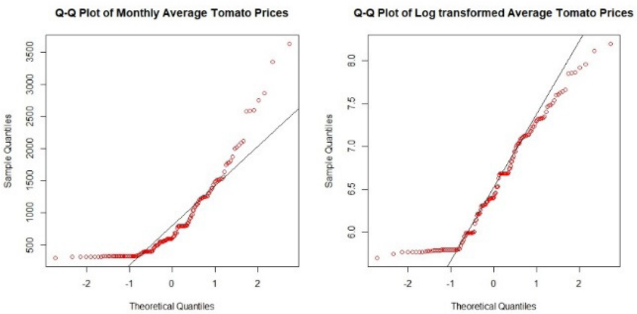

Fig. 3. Q-Q plot of original tomato price series and log transformed tomato price series

Visual inspection of the tomato price series had revealed a distinct trend and non-constant variance, indicating potential high volatility and non-stationarity. The gradual decay of autocorrelation coefficients in the Autocorrelation Function (ACF) plot in Fig. 4, also suggests that the series is non-stationary in its original form, which was further substantiated by Augmented Dickey-Fuller tests, presented in Table 2. Additionally, recurring patterns in the ACF, combined with significant lags ($12n$, $n \in 1,2$) observed in the Partial Autocorrelation Function (PACF) plot [Fig. 4 and 6], indicate the presence of seasonality with a periodicity of 12 months. On the other hand, for the first differenced series of the log-transformed tomato price series, the null hypothesis of presence of a unit root can be rejected at the 5% significance level. This indicates that the first differenced series is stationary. Since original price series is exhibiting seasonality, seasonal differencing might also benefit in modelling.

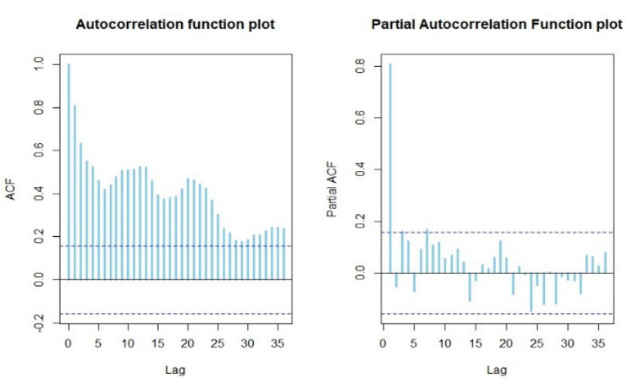

Fig. 4. Autocorrelation and Partial Autocorrelation plots of Tomato price series

Table 2. Summarised Augmented Dickey-Fuller test results for Jabalpur Tomato price series

Series	Test statistic	p-value
Tomato Price Series	-0.59	0.87
First difference of log transformed price series	-7.26	6.542×10^{-11}***

The relationship between price and arrivals appears to exhibit a hyperbolic relationship, as seen from the scatter plot in Fig. 5, which is attributed to supply demand dynamics. Thus, the data suggests a more complex, nonlinear interaction between price and arrivals, potentially involving other factors such as seasonality and long-term trends. Employing log-transformed variables, similar to the approach utilized in the second plot (Fig. 5 right), facilitates the linearization of this relationship, rendering it more amenable to modelling and forecasting.

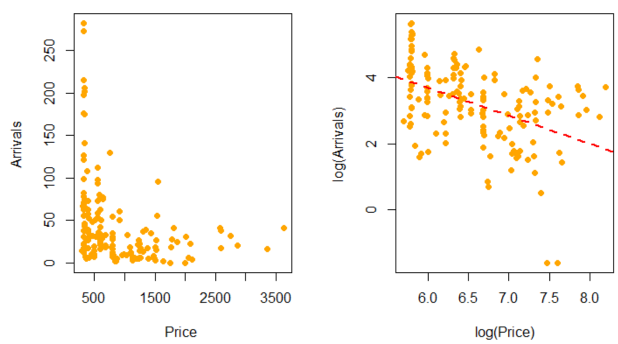

Fig. 5. Scatter plot between Arrivals and Price (left) & between log(Arrivals) and log(Price)

The Pearson correlation analysis yielded a statistically significant negative correlation ($r = -0.462$ with $p - value < 0.001$) between log-transformed price and log-transformed arrivals, though this could be spurious correlation. The optimal lag structure for the exogenous variable (log Arrivals) in the SARIMAX model was determined by fitting the model across varying lag lengths and evaluating its performance. Residual covariance analysis was conducted for each lag r, where the residual covariance matrix was computed to measure unexplained variance. Lower values of covariance indicated improved model fit. The number of lags for the log-transformed Arrivals series was determined using both HQ(n) and SC(n), as presented in Table 3. Based on these results, two lags, as suggested by HQ(n), were included in the SARIMAX

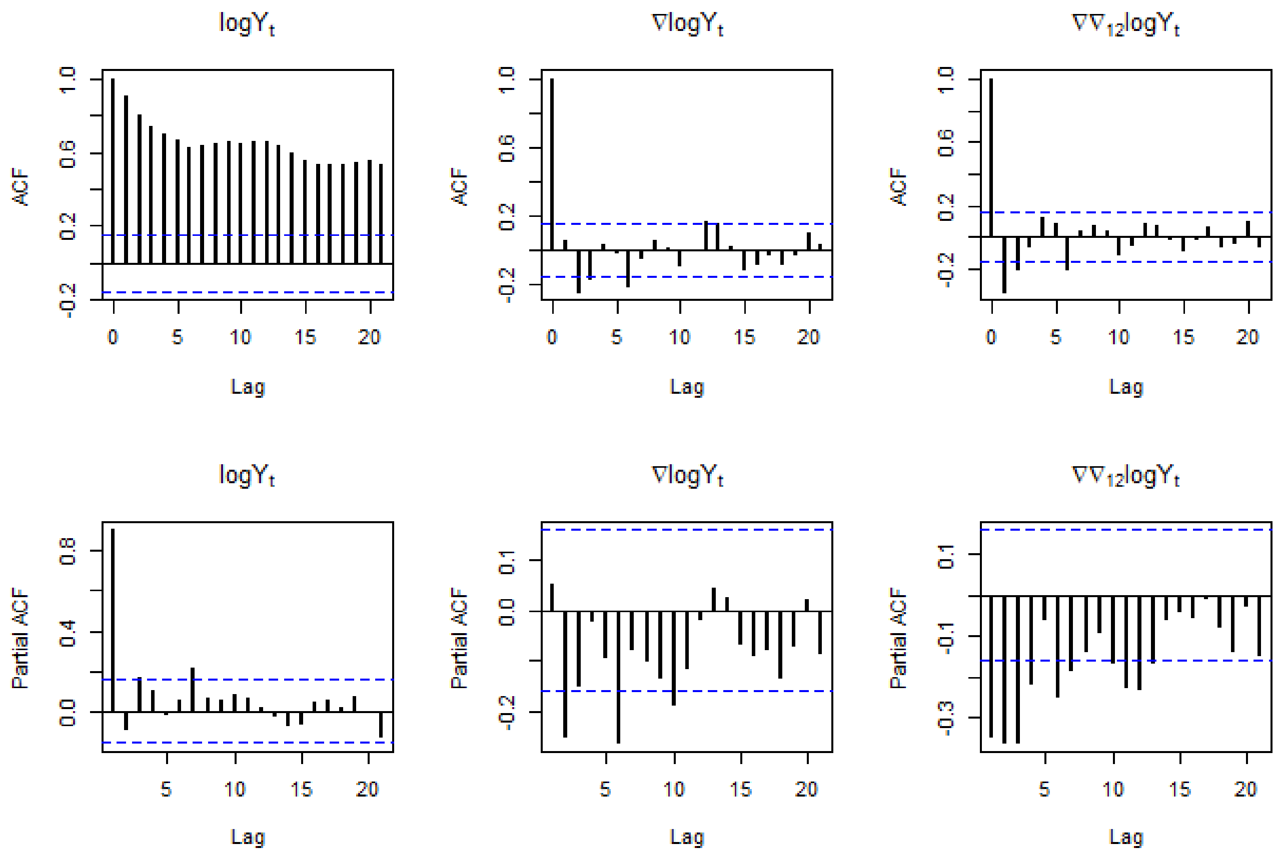

Fig. 6. Autocorrelation and Partial Autocorrelation function plots

models. Granger causality test was employed to assess the causal relationship between the first differences of log Arrivals series and the first difference of log Prices series, the results of which are presented in Table 3.

Table 3. Optimal lag determination of Exogeneous Variable & Granger Causality Test Results

Criterion	Optimal Number of lags of Exogeneous variable (X_t)	Null Hypothesis	F	p-value
HQ(n)	**2**	X_t does not Granger-cause Y_t	1.8981	0.15
SC(n)	1			

The results in Table 3 show a non-significant causal relationship between price and arrivals. The F-statistic of 1.8981 with the p-value of 0.1535, does not provide enough evidence to reject the null hypothesis. Therefore, it can be concluded that X does not Granger-cause Y. Nevertheless, aligning with the economic theory of supply and demand, where arrivals (a proxy for supply) directly influence pricing dynamics, two lags of log transformed Arrivals are included as predictors of log transformed Price in SARIMAX models to observe for any significant improvement in forecasting accuracy of prices. Tomato time series spans for a period of 157 months from June 2011 to June 2024. For model efficiency evaluation, the dataset is split into training dataset consisting of data from June 2011 to October 2021 (125 data points) and testing dataset consisting of data from November 2021 to June 2024 (32 data points). Log transformed data has been used for fitting models of appropriate orders. The appropriate orders of the models to be evaluated are determined by inspecting the ACF and PACF plots of Log transformed price series, first differenced price series and seasonally differenced first-differenced price series which are illustrated in Fig. 6. In addition to the ARIMA models, to address the seasonality component, SARIMA models of various orders have been systematically assessed.

Table 4. Model Performance Metrics for model comparison (Log-Transformed Scale)

Model	Train Data				Test Data		
	AIC	BIC	R^2	Theil's U_2	RMSE	MAPE	MAE
ARIMA(0,1,0)	11.08	13.90	0.79	1.00	0.40	4.76	0.35
ARIMA(1,1,1)	5.31	13.77	0.81	0.96	0.38	3.82	0.29
ARFIMA(1,0.492,1)	5.10	**13.58**	0.80	0.99	0.53	5.00	0.39
SARIMA(0,1,3)(0,0,1)$_{12}$	4.13	18.23	0.82	0.94	0.38	3.81	0.29
SARIMA(1,1,1)(1,0,0)$_{12}$	**3.97**	15.26	0.82	0.95	0.39	3.81	0.29
SARIMA(3,1,3)(1,0,0)$_{12}$	6.79	29.35	0.82	0.92	0.38	3.70	0.28
SARIMA(4,1,4)(1,1,1)$_{12}$	33.28	63.18	0.85	0.84	0.35	3.72	**0.27**
SARIMA(4,1,5)(1,0,0)$_{12}$	10.05	41.08	0.83	0.91	0.38	**3.70**	0.28
SARIMA(4,1,5)(1,1,1)$_{12}$	33.35	65.97	0.86	0.82	0.35	3.73	0.27
SARIMA(5,1,3)(1,1,1)$_{12}$	33.23	63.14	0.86	0.84	0.35	3.75	0.28
SARIMA(5,1,4)(1,1,1)$_{12}$	33.18	65.80	0.86	0.82	**0.35**	3.72	0.27
SARIMA(5,1,5)(0,1,1)$_{12}$	33.78	66.40	0.86	0.82	0.35	3.77	0.28
SARIMA(5,1,5)(1,1,1)$_{12}$	32.93	68.27	0.87	0.80	0.35	3.80	0.28
SARIMAX(3,1,5,2)(1,0,1)$_{12}$	11.24	47.91	0.84	0.89	0.39	4.11	0.31
SARIMAX(4,1,3,2)(1,1,1)$_{12}$	34.66	67.29	0.86	0.84	0.37	4.07	0.30
SARIMAX(5,1,5,2)(1,1,1)$_{12}$	34.48	75.26	0.87	0.79	0.37	4.06	0.30
SARIMAX(6,1,5,2)(0,1,1)$_{12}$	37.29	78.06	0.87	0.81	0.43	5.29	0.38
SARIMAX(6,1,6,2)(0,1,1)$_{12}$	38.84	82.33	**0.87**	**0.79**	0.42	5.08	0.37

* The best values amongst all criteria are in bold.

The price series under investigation exhibited signs of non-stationarity, as indicated by preliminary diagnostics such as the Augmented Dickey-Fuller (ADF) test in Table 2. However, rather than applying traditional integer differencing (d=1) to achieve stationarity, fractional differencing was also considered. Fractional differencing allows for achieving stationarity while preserving long-term dependencies which may be critical for understanding the dynamics of the price series. The fractional differencing parameter for ARFIMA (1,d,1) model is estimated to be 0.492. The estimated d value close to 0.5 indicates long-range dependence structure in the series suggesting that the price series retains substantial memory, with historical price levels over longer horizons influencing future movements. For tomato price series, such dependencies could arise from seasonal production cycles, delayed responses to supply shocks, or persistent market interventions. Amongst the several models evaluated, only a few outperforming models in terms of the minimal forecasting error metrics alongside AIC, BIC values have been outlined in the Table 4. The model selection process reveals a nuanced interplay between model complexity, the inclusion of exogenous variables, and forecasting accuracy. Upon examining the performance metrics on the log-transformed scale (Table 4), several insights emerge. Despite the ARFIMA (1, 0.4916, 1) model having the lowest BIC value (13.58) among all evaluated models, it does not justify its use, particularly with regard to out-of-sample forecasting accuracy. The weaker fit and poor out of sample forecasting performance indicated by high values of RMSE, MAPE, MAE among all evaluated models as shown in Table 4, suggests that modeling long-term dependencies through fractional differencing does not provide significant improvements in forecasting accuracy. Models such as SARIMAX(6,1,6,2)(0,1,1)$_{12}$ and SARIMAX(5,1,5,2)(1,1,1)$_{12}$ exhibit strong in-sample performance, as reflected by high R-squared values and low Theil's U_2. However, they perform poorly when applied to the test data, indicating limited generalization capability. These models exhibit relatively higher RMSE, MAE, and MAPE on the testing set as compared to their counterpart SARIMA models, suggesting over fitting to the training data.

In contrast, models like SARIMA(4,1,4)(1,1,1)$_{12}$, SARIMA(5,1,4)(1,1,1)$_{12}$, and SARIMA(3,1,4)(1,1,1)$_{12}$, which yield the lowest forecasting errors,

Table 5. Model Performance Metrics for model comparison (Original Scale)

Model	Train Data				Test Data		
	AIC	BIC	R^2	Theil's U_2	RMSE	MAPE	MAE
ARIMA(0,1,0)	5.31	13.77	0.69	0.93	720.75	25.76	479.56
ARIMA(1,1,1)	11.08	13.90	0.64	1.00	689.67	39.48	575.98
ARFIMA(1,0.492,1)	5.10	**13.58**	0.67	0.97	956.87	28.00	645.73
SARIMA(0,1,3)(0,0,1)$_{12}$	4.13	18.23	0.69	0.93	727.64	25.37	477.98
SARIMA(1,1,1)(1,0,0)$_{12}$	**3.97**	15.26	0.69	0.93	731.94	25.24	479.15
SARIMA(3,1,3)(1,0,0)$_{12}$	6.79	29.35	0.70	0.91	728.92	**24.61**	466.78
SARIMA(4,1,4)(1,1,1)$_{12}$	33.28	63.18	0.74	0.86	627.23	28.79	**455.69**
SARIMA(4,1,5)(1,0,0)$_{12}$	10.05	41.08	0.69	0.93	726.79	24.68	466.58
SARIMA(4,1,5)(1,1,1)$_{12}$	33.35	65.97	0.74	0.85	631.46	28.53	457.72
SARIMA(5,1,3)(1,1,1)$_{12}$	33.23	63.14	0.74	0.86	631.70	28.90	458.93
SARIMA(5,1,4)(1,1,1)$_{12}$	33.18	65.80	0.75	0.85	630.52	28.41	456.24
SARIMA(5,1,5)(0,1,1)$_{12}$	33.78	66.40	0.75	0.84	**623.55**	29.53	460.96
SARIMA(5,1,5)(1,1,1)$_{12}$	32.93	68.27	0.77	**0.81**	639.46	28.89	467.68
SARIMAX(3,1,5,2)(1,0,1)$_{12}$	11.24	47.91	0.73	0.89	720.33	27.85	501.56
SARIMAX(4,1,3,2)(1,1,1)$_{12}$	34.66	67.29	0.76	0.85	658.34	32.04	500.91
SARIMAX(5,1,5,2)(1,1,1)$_{12}$	34.48	75.26	**0.80**	**0.78**	660.97	32.39	503.37
SARIMAX(6,1,5,2)(0,1,1)$_{12}$	37.29	78.06	0.78	0.80	736.77	47.71	650.83
SARIMAX(6,1,6,2)(0,1,1)$_{12}$	38.84	82.33	0.79	0.78	710.61	44.94	619.73

* The best values amongst all criteria are in bold.

underscore the notion that incorporating exogenous variables does not necessarily improve predictive performance. However, the marginal improvements in accuracy observed in SARIMA(5,1,4)(1,1,1)$_{12}$ compared to the SARIMA(1,1,1)(1,0,0)$_{12}$ model do not justify the added complexity, as reflected by the higher AIC and BIC values of the former. The latter has the lowest AIC and comparable forecasting performance, demonstrating its optimality as a parsimonious yet effective model. Parsimony is a fundamental principle in time series modeling, where simpler models that adequately capture the underlying data are preferred. The chosen model, SARIMA(1,1,1)(1,0,0)$_{12}$, exhibits satisfactory forecast error metrics, including RMSE (0.385), MAE (0.286), and MAPE (3.812). When considering model performance on the original scale presented in Table 5, MAPE, a metric robust to transformations reveals minimal differences in accuracy between SARIMA(1,1,1)(1,0,0)$_{12}$ (MAPE = 25.24) and SARIMA(3,1,3)(1,0,0)$_{12}$ (MAPE = 24.61). The latter, with a more complex structure, has higher BIC (29.35) values, suggesting that SARIMA(1,1,1)(1,0,0)$_{12}$ provides a parsimonious and robust choice for forecasting without sacrificing predictive accuracy. This consistency of its performance across both the log-transformed and original scales reinforces the idea that simpler models, such as SARIMA(1,1,1)(1,0,0)$_{12}$, are sufficient for forecasting tasks.

The performance of the SARIMA(1,1,1)(1,0,0)$_{12}$ model for forecasting tomato prices, as visualized in the Fig. 7, highlights its strengths and areas for improvement. It demonstrates its capability for short-duration forecasting by closely following the real trajectory of tomato prices. For short-term forecasting, the model accurately tracks the real dynamics of price changes, showcasing its strength in scenarios where abrupt shifts are limited. During the training period, the model's fitted values closely align with the actual time series, effectively capturing the seasonal and trend components. However, as the forecasting horizon extends, the model struggles to capture extreme price spikes and troughs, likely due to its inability to account for volatility clustering or heteroscedasticity, which are often observed in 'TOP' commodities price data. In the test period, the divergence between actual and predicted values becomes more pronounced, with the predicted values displaying a relatively flat trajectory appearing relatively smoother and less reactive to abrupt changes in the actual series, indicating the model's difficulty in extrapolating highly dynamic patterns.

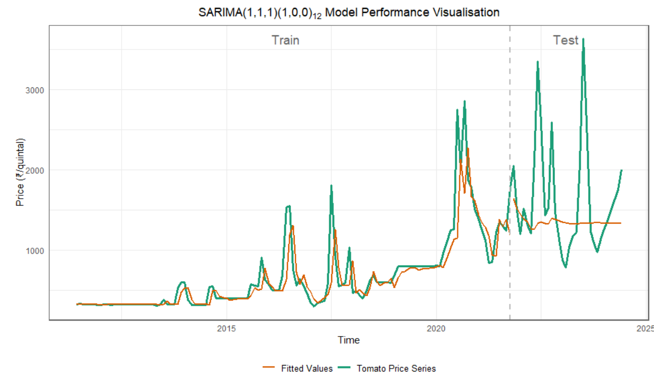

Fig. 7. SARIMA(1,1,1)(1,0,0)$_{12}$ Model Performance Visualisation

Although the SARIMA (1,1,1)(1,0,0)$_{12}$ model adeptly manages autocorrelations as is evident through Ljung Box test results in Table 6, it does not explicitly account for variations in volatility over time. To address limitation of not capturing volatility clustering, integrating ARCH/GARCH models could enhance the SARIMA(1,1,1)(1,0,0)$_{12}$ model's performance. In the analysis of the residuals, we observe the phenomenon of volatility clustering, characterized by consecutive periods of elevated and diminished variance. Furthermore, the ARCH LM test results in Table 6, reveal the presence of conditional heteroscedasticity in the residuals, as evidenced by significant test statistics. These findings suggest that conditional heteroscedasticity remains an unexplained pattern in the residuals.

Table 6. Results of the Diagnostic tests on the SARIMA(1,1,1)(1,0,0)$_{12}$residuals

Diagnostic Test	χ^2 statistic	df	p-value
Box-Ljung Test	32.12	24	0.124
ARCH LM	5.31	1	0.02 **
	6.56	2	0.04**

Since the ARCH LM test reveals significant ARCH effects in the residuals of the SARIMA (1, 1, 1)(1, 0, 0)$_{12}$ model, applying ARCH/GARCH models is justified to address the time-varying volatility (conditional heteroscedasticity) that SARIMA fails to capture. By modeling the conditional variance, ARCH/GARCH models ensures more reliable confidence intervals and rectifies inefficiencies in the SARIMA(1,1,1)(1,0,0)$_{12}$ model's estimates, making it crucial for more accurate predictions in contexts with high volatility. To effectively capture the volatility clustering observed in the data, several ARCH and GARCH models were fitted to the residuals. Model selection was based on the Akaike Information Criterion (AIC), with the results presented below in Table 7. The standard GARCH(1,1) model, which exhibited the lowest AIC, was selected as the most parsimonious and effective model in capturing the volatility dynamics of the tomato price series. Fig. 8 presents the volatility forecasts derived from the GARCH(1,1) model against the actual variance of the SARIMA(1,1,1)(1,0,0)$_{12}$ model residuals.

Table 7. AIC Values for Various ARCH and GARCH Model Specifications

Model	AIC
ARCH(1)	0.08
ARCH(2)	0.05
sGARCH(1,1)	**0.02**
sGARCH(2,1)	0.04
sGARCH(2,2)	0.04

The GARCH(1,1) model captures volatility clustering in the SARIMA(1,1,1)(1,0,0)$_{12}$ residuals. The fit (predicted conditional variance) of the GARCH(1,1) model appears to effectively track the time-varying nature of the volatility, suggesting that the model is well-suited for capturing the persistence of volatility shocks in the data as shown in Fig. 8.

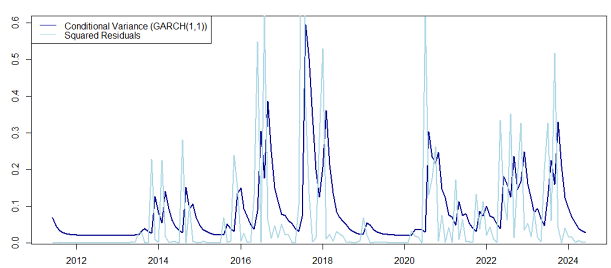

Fig. 8. Volatility of SARIMA (1, 1, 1)(1, 0, 0)$_{12}$ residuals captured by GARCH (1,1)

The hybrid SARIMA(1,1,1)(1,0,0)$_{12}$–GARCH(1,1) model, with lower computed RMSE (0.23), MAPE (3.17) and MAE (0.24), shows considerable improvement compared to the SARIMA(1,1,1)(1,0,0)$_{12}$ model alone. The Table 8 summarizes the parameter estimates for the hybrid SARIMA(1, 1, 1)(1, 0, 0)$_{12}$ - GARCH(1,1) model. The mean and the variance equations for the model whose coefficients were estimated in Table 8 is expressed as follows,

$$y_t = \zeta_t + 0.97\zeta_{t-1} + 0.09y_{t-1} + 0.90y_{t-2} + 0.18y_{t-12} - 0.02y_{t-13} - 0.16y_{t-14}$$

$$\zeta_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = 0.009102 + 0.450462\zeta_{t-1}^2 + 0.548538\sigma_{t-1}^2$$

where, $y_t = log(Y_t)$ (logarithm of the original tomato price series Y_t), σ_t^2 is the conditional variance of ζ_t and ϵ_t is white noise.

Table 8. Parameter Estimation of the hybrid SARIMA (1, 1, 1) (1, 0, 0)$_{12}$ - GARCH(1,1) model

Model Specification	**SARIMA (1, 1, 1) (1, 0, 0)$_{12}$**			
Parameter	**Estimate**	**Std. Error**	**t-statistic**	**p-Value**
AR1	-0.91	0.06	-15.49	0.00***
MA1	0.97	0.03	28.20	0.00***
SAR1	0.18	0.08	2.21	0.03**
Constant	0.01	0.03	0.43	0.66
Model Fit Statistic	Value			
Variance (σ^2)	0.07			
Log-likelihood	-12.01			
AIC	32.21			
Model Specification	**sGARCH(1,1)**			
Parameter	**Estimate**	**Std. Error**	**t-Value**	**p-Value**
Omega (ω)	0.01	0.01	0.66	0.51
Alpha1 (α1)	0.45	0.19	2.42	0.02**
Beta1 (β1)	0.55	0.23	2.35	0.02**
Distribution	Normal			
Model Fit Statistic	Value			
Log-likelihood	1.53			
AIC	0.02			
BIC	0.08			

The diagnostic analysis of the hybrid SARIMA (1, 1, 1)(1, 0, 0)$_{12}$ - GARCH(1,1) model residuals in Fig. 9, suggests that the model is well-specified and adequately captures both the mean and volatility dynamics of the data. The standardized residuals fluctuate randomly around zero without visible patterns, and the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots indicate no significant autocorrelations at all lags, as the values remain within the 95% confidence bounds as observed in Fig. 9. The Ljung-Box test results in Table 9 provides insufficient evidence to reject the null hypothesis of no autocorrelation. The Ljung Box p-value plot in Fig. 9, further supports this, reinforcing the conclusion of no autocorrelation in residuals. The Weighted ARCH-LM tests (Li & Mak, 1994) in Table 9, provides no evidence of remaining ARCH effects, indicating that the GARCH(1,1) component has effectively captured the time-varying volatility. However, the residuals exhibit significant departures from normality, as evidenced by the both Jarque-Bera & Shapiro Wilk test results in Table 9, which strongly reject the null hypothesis of normality. This finding aligns with the histogram of standardized residuals in Fig. 9, which reveals excess kurtosis and heavy tails. These results suggest that while the hybrid SARIMA-GARCH model successfully captures serial correlation and conditional heteroscedasticity, it fails to account for the non-normality in the residuals. Overall, the model performs well in terms of addressing autocorrelation and volatility clustering and is suitable for forecasting purposes, although improvements could be made to account for residual non-normality.

Table 9. Results of Residual Diagnostic tests of the Hybrid model

Diagnostic Tests	χ^2 **statistic**	**p-value**
Box-Ljung Test	28.45	0.24
Weighted ARCH-LM	0.13 (Lag = 3)	0.71
	0.96 (Lag = 7)	0.91
Jarque-Bera	211.78	$<2.2\times10^{-16}$***
Shapiro-Wilk	0.84	9.721×10^{-12}***

The forecast visualised in Fig. 10 presents the projected tomato prices in the Jabalpur district for a 12-month horizon, from July 2024 to June 2025. In addition to mean forecasts obtained from the hybrid SARIMA (1,1,1)(1,0,0)$_{12}$-GARCH(1,1) model, the forecast includes the 95% and 90% confidence intervals. As can be observed in Fig. 10, the confidence intervals for the forecasts exhibit notable asymmetry, with the upper bounds consistently larger than the lower bounds. This asymmetry can be attributed to the presence of skewness in the distribution of the innovations, indicating that positive shocks may have a more significant and sustained impact on volatility than negative shocks. Additionally, the leverage like effect, where positive shocks lead to greater volatility, likely

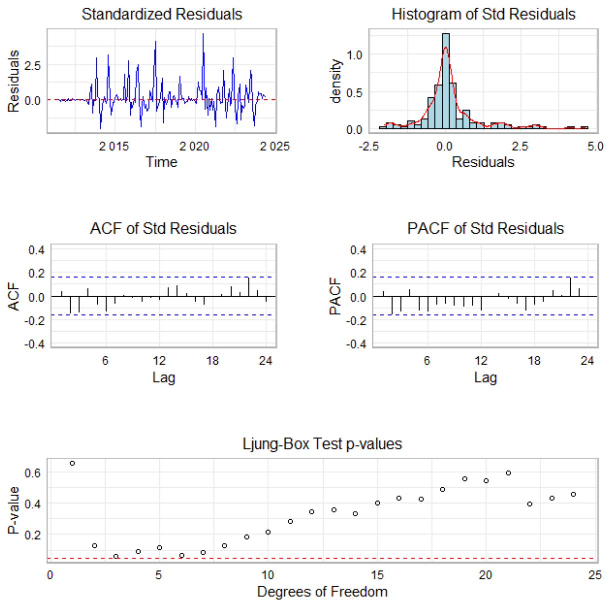

Fig. 9. Residual Diagnostic tests of hybrid SARIMA(1, 1, 1)(1, 0, 0)$_{12}$ - GARCH(1,1) model

contributes to this pattern. The non-normal distribution of residuals in the hybrid SARIMA (1, 1, 1) (1, 0, 0)$_{12}$ - GARCH (1,1) model further intensifies this asymmetry, as fat tailed or skewed residuals increase the likelihood of larger positive deviations, thus widening the upper bounds of the confidence intervals.

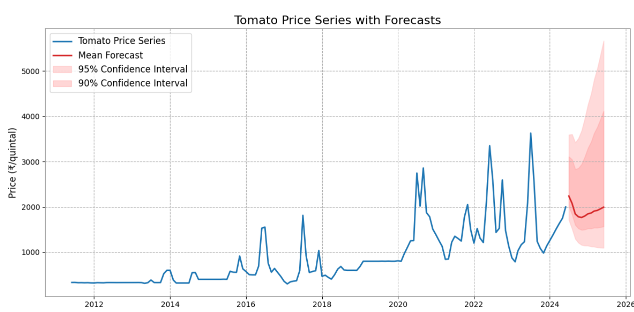

Fig. 10. Monthly Tomato Price Series with Forecasts

4. CONCLUSION

This study demonstrates that the hybrid SARIMA(1,1,1)(1,0,0)$_{12}$–GARCH(1,1) model is a robust and parsimonious tool for short-term tomato price forecasting in the Jabalpur district. By integrating SARIMA's capacity to model seasonal structure with GARCH's ability to capture volatility clustering, the framework effectively addresses both deterministic and stochastic components of price behavior. Although the price series exhibited moderate long-memory (d = 0.4916), the ARFIMA model did not outperform SARIMA(1,1,1)(1,0,0)$_{12}$, reinforcing the practical adequacy of the simpler model. Volatility diagnostics confirmed the suitability of GARCH (1,1) for modelling heteroskedasticity. Attempts to enhance forecast performance using two lags of log-transformed arrivals as exogeneous variables in a SARIMAX model yielded limited gains, underscoring the value of model parsimony. The findings offer actionable insights for policymakers and market participants by enabling more informed decision-making related to production, storage, and pricing strategies. Moreover, the proposed framework maybe adaptable to other agricultural commodities exhibiting similar seasonal and volatile characteristics. Tomato prices exhibit significant fluctuations across different months, with prices occasionally increasing manifold times compared to the peak harvest period prices. If farmers strategically plan their cultivation area, sowing dates, and sales by incorporating forecasts derived from hybrid model, they could potentially increase their earnings by three to four times. However, the accuracy of these earnings predictions relies on the forecast model's ability to provide reliable forecasts within a 90% confidence interval (Boateng *et al.,* 2017). Accurate price forecasts offer critical insights for traders, enabling them to decide whether to import tomatoes from other markets to capitalize on higher local prices in comparison to international prices (Adanacioglu and Yercan, 2012). Additionally, tomato prices forecasts can serve as a valuable tool for both tomato growers in Jabalpur and policymakers, aiding in informed decision-making for production, marketing strategies, and market interventions aimed at stabilizing prices and ensuring food security. Future research can focus on enhancement of the forecasting framework by incorporation of other exogenous variables such as rainfall, temperature, input costs, and export-import data apart from arrivals (in mandis) and assess for any improvement in forecast accuracy for tomato prices and also assess the hybrid model's applicability in other districts or states to generalize findings across different agricultural zones.

REFERENCES

Adanacioglu H. and Yercan M. (2012). An analysis of tomato prices at wholesale level in Turkey: An application of SARIMA model. *Custos e Agronegocio*, **8**, 52-75.

Badal, P.S., Kamalvanshi, V., Goyal, A., Kumar, P., and Mondal, B. (2022). Forecasting potato prices: Application of ARIMA model. *Economic Affairs*, **67**(4), 491-496.

Boateng, F.O., John, A.M., Martin, A., Lot, O., and Paul, D. (2017). Modeling of tomato prices in Ashanti region, Ghana, using seasonal autoregressive integrated moving average model. *British Journal of Mathematics & Computer Science*, **20**(1), 1-13.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G.E.P., and Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Casolaro, A., Capone, V., Iannuzzo, G., and Camastra, F. (2023). *Deep Learning for Time Series Forecasting: Advances and Open Problems*. Information, **14**(11), 598.

Dickey, D.A., & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**(366), 427-431. https://doi.org/10.2307/2286348

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**(4), 987-1007. https://doi.org/10.2307/1912773

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **222**(594-604), 309-368. https://doi.org/10.1098/rsta.1922.0009

Government of India, Ministry of Finance, Department of Economic Affairs. (2025). *Economic Survey 2024–25 Chapter 4 – Prices and Inflation: Understanding the Dynamics.* https://www.indiabudget.gov.in/economicsurvey/doc/eschapter/echap04.pdf

Jain, A., Marvaniya, S., Godbole, S., and Munigala, V. (2022). Towards context-based model selection for improved crop price forecasting. *In Proceedings of the ACM Conference*.

Kamu, A., Ahmed, A., and Yusoff, R. (2010). Forecasting cocoa bean prices using univariate time series models. *Journal of Arts Science & Commerce*, **1,** 71.

Kitworawut, P., and Rungreunganun, V. (2019). Corn price modeling and forecasting using Box-Jenkins model. *Applied Science and Engineering Progress*.

Knotters, M., Brus, D.J., and van der Ploeg, M.J. (2010). A disposition of interpolation techniques. Wageningen UR — Alterra / Report (chapter on state-space approach).

Kumar, T.L., and Mohan, S. (2011). Holt-Winters exponential smoothing and seasonal ARIMA time-series technique for forecasting of onion price in Bangalore market.

Priya, L.K., Suresh, P., Yeturu, N., Bhargavi, N., Mathew, M., Mansingh, P.J., and Nisha, A. (2024). Price behaviour and price forecasting of tomato in Koyambedu Market, Tamil Nadu. *International Journal of Religion,* **5**(10), 2954-2962.

Li, W.K., and Mak, T.K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity. *Journal of Time Series Analysis*, **15**(6), 627-636.

Limpert, E., Stahel, W.A., and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, **51**(5), 341-352.

Ljung, G. M., and Box, G.E.P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, **65**(2), 297-303.

National Bank for Agriculture and Rural Development (NABARD). (2023). *Volatility in TOP and role of agri value chain* (Ecofocus, Issue No. 02, June). https://www.nabard.org/auth/writereaddata/tender/pub_270623091211322.pdf

Patra M.D., J John and George, A.T. (2023), Recent Regime Reversal in Inflation: The Indian Experience, *RBI Bulletin, April*.

Paul, R., Rana, S., and Saxena, R. (2016). Effectiveness of price forecasting techniques for capturing asymmetric volatility for onion in selected markets of Delhi. *The Indian Journal of Agricultural Sciences*, ***86***, 303-309.

Pramanik, R., Alam, M.W., Ray, M., Nayak, H., Kumar, R.R., and Chaturvedi, K.K. (2023). A comparative study of advanced forecasting models on volatile time-series price data. *Journal of the Indian Society of Agricultural Statistics*, **77**(2), 155-170.

Pratap, B., Ranjan, A., and Kishore, V. (2021). Forecasting food inflation using news-based sentiment indicators. *Reserve Bank of India Occasional Papers*, **42**(2). Rajpoot, K., Singla, S., Singh, A., & Shekhar, S. (2021). Impact of COVID-19 lockdown on prices of potato and onion in metropolitan cities of India. *Journal of Agribusiness in Developing and Emerging Economies*.

Reddy, A.A., Raju, S.S., Suresh, A., and Kumar, P. (2018). Analysis of pearl millet market structure and value chain in India. *Journal of Agribusiness in Developing and Emerging Economies*, ***8***, 00-00.

Sanjeev, Pushpa, Vikram, Preeti, and Pawan Kumar. (2025). Comparative Study of ARIMA, SARIMA and Hybrid (ARIMA + ANN and SARIMA + ANN) Models for Wholesale Monthly Average Price of Tomato and Onion. *Journal of the Indian Society of Agricultural Statistics*, ***77***(03), 275-284.

Sarkar, S., Das, S., Dutta, S., Bera, B., Bera, M., Roy, D., Middya, M., and Choudhury, M. (2024). Economic analysis of the behavior of prices and market arrivals of tomatoes: A case study of West Bengal, India. *Agricultural & Rural Studies*, **2***,* 1-12.

Singh, U., Singh, R.B., and Rajendra, B.L. (2015). Econometric simulation model of dal industries for forecasting and policy decision. *Journal of The Indian Society of Agricultural Statistics*, **69**(3), 301-305.

Sonvanee, O., and Bhargav, P. (2024). Forecasting tomato price and arrival patterns in Krishi Upaj Mandis, Rajnandgaon, Chhattisgarh using ARIMA models. *Asian Journal of Agricultural Extension Economics & Sociology*, **42**, 25-36.

Vinay, H.T., Basak, P., Ghosh, A., Ojha, S., and Sahu, C.R. (2024). Forecasting of tomato price in Karnataka using BATS model. *Journal of the Indian Society of Agricultural Statistics*, **78**(2), 107-113.